

# The Uncertainty (Validity and Reliability) of Three Electrothermometers in Therapeutic Modality Research

Lisa S. Jutte; Kenneth L. Knight; Blaine C. Long; Jeremy R. Hawkins;  
Shane S. Schulthies; Ethan B. Dalley

Brigham Young University, Provo, UT

*Lisa S. Jutte, PhD, ATC, contributed to conception and design; acquisition and analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Kenneth L. Knight, PhD, ATC, FACSMT, contributed to conception and design; analysis and interpretation of the data; and drafting, critical revision, and final approval of the article. Blaine C. Long, MS, ATC, contributed to conception and design and drafting, critical revision, and final approval of the article. Jeremy R. Hawkins, MS, ATC, CSCS, contributed to analysis and interpretation of the data and critical revision and final approval of the article. Shane S. Schulthies, PhD, ATC, PT, contributed to analysis and interpretation of the data and drafting, critical revision, and final approval of the article. Ethan B. Dalley contributed to acquisition of the data and drafting, critical revision, and final approval of the article.*

*Address correspondence to Lisa S. Jutte, PhD, ATC, School of Physical Education, Sport, and Exercise Science, Ball State University, HP 222 H, Muncie, IN 47306. Address e-mail to lsjutte@bsu.edu.*

**Context:** Data from electrothermometers are used to determine therapeutic modality usage, but the value of experimental results is only as good as the data collected.

**Objective:** To determine the reliability and validity of 3 electrothermometers from 2 manufacturers.

**Design:** A  $3 \times 4 \times 17$  factorial with repeated measures on 2 factors. Independent variables were trial (1, 2, 3), thermometer (mercury thermometer, Iso-Thermex calibrated from  $-50^{\circ}\text{C}$  to  $50^{\circ}\text{C}$ , Iso-Thermex calibrated from  $-20^{\circ}\text{C}$  to  $80^{\circ}\text{C}$ , and Datalogger), and time (17).

**Setting:** Human Performance Research Center.

**Intervention(s):** Eighteen thermocouples were inserted through the wall of a foamed polystyrene cooler, and 6 were connected to each of the 3 electrothermometers. The cooler was positioned on a stir plate and filled with room-temperature water ( $18.4^{\circ}\text{C}$ ). A mercury thermometer was immersed into the water bath. Measurements of the water bath were taken every 10 seconds for three 3-minute trials.

**Main Outcome Measure(s):** The temperature variability of 3 electrothermometers was taken from a calibrated mercury thermometer.

**Results:** The Iso-Thermex electrothermometers did not differ statistically from each other in uncertainty (validity error  $\pm$  reliability error =  $0.06^{\circ}\text{C} \pm 0.03^{\circ}\text{C} \pm 0.03^{\circ}\text{C} \pm 0.02^{\circ}\text{C}$ ,  $P < .05$ ), but both differed from the Datalogger ( $0.64^{\circ}\text{C} \pm 0.20^{\circ}\text{C}$ ,  $P < .05$ ). The Datalogger temperature was consistently higher than the mercury thermometer temperature.

**Conclusions:** The Iso-Thermex electrothermometers were more stable than the Datalogger, and values were within the published uncertainty ( $\pm 0.1^{\circ}\text{C}$ ) when used with PT-6 thermocouples. The Datalogger we used had an uncertainty of measurement greater than that indicated in the user's manual ( $\sim \pm 0.52^{\circ}\text{C}$ ). Uncertainty of  $\pm 0.84^{\circ}\text{C}$  can significantly influence the interpretation of results when intramuscular temperature changes are usually less than  $5^{\circ}\text{C}$ .

**Key Words:** accuracy, repeatability, reproducibility, measurements

One indicator of the efficacy of a thermal therapeutic modality is its ability to change temperature in specific tissues. Measuring tissue temperature, therefore, is an important element in therapeutic modality research. The value of experimental results is only as good as the data collected. Both the Iso-Thermex (Columbus Instruments, Columbus, OH) and Datalogger (model MMS 3000-T6V4; Commtest Instruments Ltd, Christchurch, New Zealand) electrothermometers have been used for measuring tissue temperature and have helped to advance our understanding of thermal and cryotherapeutic modalities.

Most therapeutic modality investigators using these electrothermometers have either reported the manufacturer's claims of their instrument's accuracy<sup>1–4</sup> or failed to report the reliability and validity of their instruments.<sup>5–12</sup> We know of

no independent assessment of these two electrothermometers. Our research question was, How reliable and valid are the Iso-Thermex and Datalogger?

## METHODS

### Experimental Design

A  $3 \times 4 \times 17$  repeated-measures design guided this study. The independent variables were trial (1, 2, and 3), thermometer (mercury thermometer, Iso-Thermex calibrated from  $-50^{\circ}\text{C}$  to  $50^{\circ}\text{C}$ , Iso-Thermex calibrated from  $-20^{\circ}\text{C}$  to  $80^{\circ}\text{C}$ , and Datalogger), and time (every 10 seconds for 3 minutes,  $n = 17$ ). The dependent variables were temperature and absolute temperature differences from a mercury thermometer.



**Figure 1.** Datalogger, Iso-Thermex –20:80, and Iso-Thermex –50:50 electrothermometers stacked (top to bottom) in center of picture. Computer for Iso-Thermex –20:80 in background; computer for Iso-Thermex –50:50 in foreground right. Foamed polystyrene cooler in foreground left.

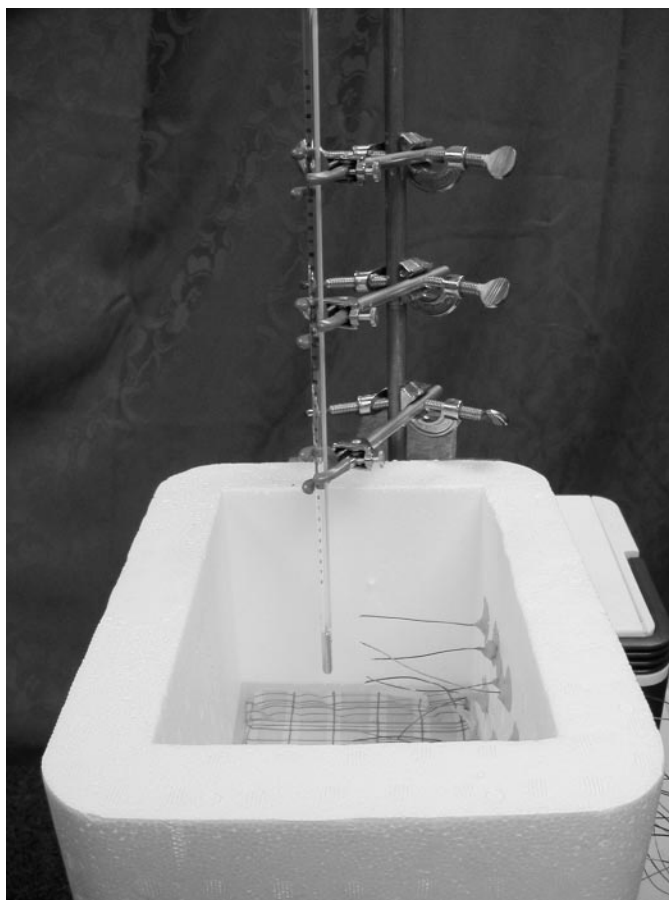
## Instruments

We interfaced 18 PT-6 thermocouples (Physitemp Instruments Inc, Clifton, NJ) with 3 electrothermometers (6 thermocouples per machine): a 16-channel Iso-Thermex with a measurable temperature range of  $-50^{\circ}\text{C}$  to  $50^{\circ}\text{C}$  (Iso–50:50), a second 16-channel Iso-Thermex with a measurable temperature range of  $-20^{\circ}\text{C}$  to  $80^{\circ}\text{C}$  (Iso–20:80), and a 6-channel Datalogger with a measurable temperature range of  $-250^{\circ}\text{C}$  to  $350^{\circ}\text{C}$  (Figure 1). A calibrated mercury thermometer (model 15-059-18; Fisher Scientific International Inc, Hampton, NH; National Institute of Standards and Technology<sup>13</sup> traceable) graded at  $0.1^{\circ}\text{C}$  was used to monitor water bath temperature. We circulated the water bath with a stirrer (model 103, Corning PC, Corning, NY) and magnetic stir bar.

## Procedures

Eighteen thermocouples, 3 rows of 6, were inserted through the wall of a  $23 \times 15 \times 19\text{-cm}$  foamed polystyrene cooler and secured with silicone polymer (Figure 2). The thermocouples extended 10 cm into the cooler. The bottom row was 4 cm from the cooler bottom, and rows were 3 cm apart from each other. Thermocouples within a row were also 3 cm apart from each other. The cooler was filled to within 3 cm of the top with water ( $18.4^{\circ}\text{C}$ ) that had sat in the room for 24 hours before the study. Water temperature gradients were minimized with a magnetic stir bar spinning in the bottom of the cooler. The mercury thermometer was immersed into the water bath approximately 3 cm from the bottom.

We interfaced 6 thermocouples with each of the 3 electrothermometers, using all 6 channels of the Datalogger and 6 randomly selected channels in each Iso-Thermex unit. Data collection consisted of 3 trials. After the first and second trials,



**Figure 2.** Mercury thermometer positioned in the cooler.

the 3 sets of 6 thermocouples were rotated among electrothermometers. Electrothermometers were started simultaneously, and temperature was recorded every 10 seconds for 3 minutes. The same investigator read the mercury thermometer every 10 seconds throughout each trial.

## Statistical Analysis

We computed the mean and standard deviation of the 306 measures (17 samples for each of 18 thermocouples) for each machine. The standard deviation served as our measure of reliability. We then compared the reliability (standard deviation) among electrothermometers using a modified Levenes equal variance test<sup>14</sup> and pairwise F tests.<sup>14</sup>

Validity was computed as the absolute difference between the individual measurements of the mercury thermometer and each electrothermometer. These individual differences then became the dependent variable in a  $3 \times 17$  repeated-measures analysis of variance,<sup>15</sup> with electrothermometer and time as main factors. Thermocouple was used as a repeat measure to ensure that the differences were between machines and not thermocouples. When appropriate, we computed Scheffé multiple comparison tests to determine the location of statistical differences among individual machines. Results were considered statistically significant at an alpha level of 0.05.

## RESULTS

Each Iso-Thermex unit was more valid ( $F_{2,34} = 151.51$ ,  $P = .000001$ ) and more reliable (Iso –50:50:  $F_{305,305} = 36.35$ ,

**Table 1. Means (Validity) and Standard Deviations of Absolute Temperature Differences for 3 Electrothermometers\***

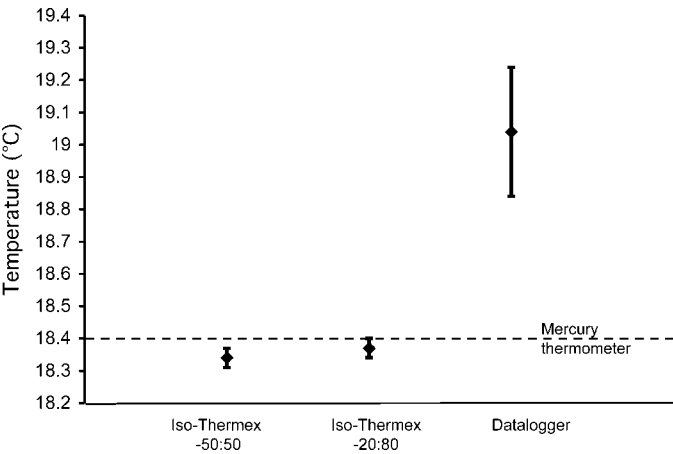
Machine	Mean (Validity)	SD
Iso-Thermex -50:50	0.06†	±0.03
Iso-Thermex -20:80	0.03†	±0.02
Datalogger	0.64†	±0.21

\*Iso-Thermex -50:50 indicates temperature range of -50°C to 50°C, and Iso-Thermex -20:80, temperature range of -20°C to 80°C.  
†Iso-Thermex -50:50 and Iso-Thermex -20:80 < Datalogger ( $P < .05$ ).

**Table 2. Means and Standard Deviations (Reliability) of Temperature Measurements for 3 Electrothermometers\***

Machine	Mean	SD (Reliability)
Iso-Thermex -50:50	18.34	±0.03†
Iso-Thermex -20:80	18.37	±0.03‡
Datalogger	19.04	±0.20‡

\*Iso-Thermex -50:50 indicates temperature range of -50°C to 50°C, and Iso-Thermex -20:80, temperature range of -20°C to 80°C.  
†Iso-Thermex -50:50 < Datalogger ( $P < .05$ ).  
‡Iso-Thermex -20:80 < Datalogger ( $P < .05$ ).



**Figure 3. Mean temperatures and standard deviations for each electrothermometer.**

$P = .00001$ ; Iso -20:80:  $F_{305,305} = 49.97$ ,  $P = .00001$ ) than the Datalogger, but they did not differ from each other (Tables 1 and 2, Figure 3). The difference between the Datalogger and the mercury thermometer was 10–20 times more than the difference between the Iso-Thermex units and the mercury thermometer (see Table 1). The Iso-Thermex -20:80 underestimated the mercury thermometer by 0.03°C; the Iso-Thermex -50:50 underestimated the mercury thermometer by 0.06°C; and the Datalogger overestimated the mercury thermometer by 0.64°C (see Table 1, Figure 3). The reliability of the Iso-Thermex units was not different, but both were more reliable than the Datalogger (see Table 2).

**DISCUSSION**

Terms used to describe how closely or consistently repeated measurements describe phenomena include *accuracy*, *conformance*, *reliability*, *validity*, *error*, *repeatability*, *reproducibility*, and *uncertainty*. Scientists use the terms *reliability* and

*validity*,<sup>16</sup> whereas manufacturers of instruments typically use the term *accuracy*.<sup>17,18</sup> However, the National Institute of Standards and Technology and those who calibrate thermometers to its specifications,<sup>13,19</sup> use the words *accuracy*, *conformance*, *error*, *repeatability*, *reproducibility*, and *uncertainty*. These terms are sometimes used synonymously and incorrectly, most likely because of a misunderstanding of their meanings. They are defined as follows:

- *Reliability* is the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials.<sup>20</sup>
- *Validity* is the extent to which a situation as observed reflects the true situation, or the degree to which data or results of a study are correct or true.<sup>21</sup>
- *Accuracy* is the freedom from mistake or error—correctness.<sup>20</sup> This includes freedom from both validity and reliability errors. Ironically, manufacturers use *accuracy* as a quantitative term despite the fact that the National Institute of Standards and Technology defines *accuracy* as a qualitative and not a quantitative term.
- *Conformance* and *conformity* are actions that are in accordance with some specified standard or authority.<sup>20</sup> This would indicate validity.
- *Error* is the difference between an observed or calculated value and a true value<sup>20</sup>—the amount of deviation from a standard or specification,<sup>20</sup> composed of random error and systematic error.<sup>15</sup> Random error is equivalent to error that would affect reliability, whereas systematic error contributes to validity.
- *Repeatability* is the closeness of agreement between the results of successive measurements carried out under the same conditions of measurement.<sup>15</sup>
- *Reproducibility* is the closeness of agreement between the results of measurements carried out under changed conditions of measurement.<sup>15</sup>
- *Uncertainty* is a value, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurement,<sup>22</sup> composed of uncertainty from both random and systematic error.<sup>15</sup> Again, random error contributes to reliability, whereas systematic error contributes to validity.

Neither Columbus Instruments nor Commtest Instruments reports the validity or reliability of their instruments.<sup>18,23</sup> Both report accuracy, which is inappropriate because they are assigning a quantitative value to a qualitative term. What they really mean to report is uncertainty (M. Grisby, unpublished data, 2005). Uncertainty comprises both random error (reliability) and systematic error (validity).<sup>15</sup> To calculate uncertainty based on this definition, we must add both our reliability and validity measurements (Table 3).

Our uncertainty measurements for both Iso-Thermex units were less (ie, they had smaller boundaries of error [conceptually similar to confidence intervals] than their manufacturer's specifications), whereas the Datalogger had greater uncertainty than its manufacturer's claims (see Table 3). Columbus Instruments reports the Iso-Thermex is accurate to within  $\pm 0.1^\circ\text{C}$ , which is more uncertain than our calculated uncertainty of  $\pm 0.09^\circ\text{C}$  for the Iso-Thermex -50:50 and  $\pm 0.06^\circ\text{C}$  for the Iso-Thermex -20:80. Commtest Instruments reports the Datalogger accuracy to be  $\pm (0.5 + [0.001 \times \text{temperature measurement}])$ ,<sup>18</sup> which would be  $\pm 0.52^\circ\text{C}$  for our study. This is less uncertain than our calculated uncertainty of  $\pm 0.84^\circ\text{C}$ .

The systematic error (validity) and random error (reliability)



**Table 3. Calculated Uncertainty and Reported Accuracy of Each Electrothermometer\***

Machine	Our Calculated Uncertainty†	Manufacturer's Reported Accuracy
Iso-Thermex –50:50	±0.09	±0.10
Iso-Thermex –20:80	±0.06	±0.10
Datalogger	±0.84	±0.52‡

\*Iso-Thermex –50:50 indicates temperature range of –50°C to 50°C, and Iso-Thermex –20:80, temperature range of –20°C to 80°C.

†Calculated uncertainty = validity + reliability.

‡Commtest Instruments Ltd does not report a specific value but offers the following formula:  $\pm (0.5 + [0.001 \times \text{temperature measurement}])$ .

measures almost equally contributed to our calculated uncertainty for the Iso-Thermex units (see Tables 1 and 2). The Datalogger's uncertainty, however, was composed mostly of systematic error (validity).

The uncertainty (combination of reliability and validity) of the Iso-Thermex and Datalogger electrothermometers is vital in interpreting the results of the therapeutic modality researchers using them. Scientists and clinicians should be aware that conclusions based on Datalogger data may be less accurate than those generated by an Iso-Thermex. Not only did the Datalogger fail to meet manufacturer's claims for accuracy, but the manufacturer's claims were more uncertain than those of the Iso-Thermex (0.52°C versus 0.1°C for the temperature in our study). This may be especially relevant when the reported difference between experimental conditions is within 1°C.

Because we evaluated a limited number of units at a single temperature, our results may not apply to all units of a similar model and brand of electrothermometers, nor to other temperatures. Therefore, the more important conclusion of this study is not the actual reliability and validity values for our specific machines. Our observations demonstrate the need for individual assessment and reporting of reliability and validity for individual machines. Scientists examining tissue temperature should know the manufacturer's claims of uncertainty, test their own equipment, and report the reliability and validity of their instruments, so that others are aware of the uncertainty of the measurements. Knowing the uncertainty of research results helps consumers of the literature make more correct interpretations of those results and may help to explain differences found among studies.

## REFERENCES

- Burr PO, Demchak TJ, Cordova ML, Ingersoll CD, Stone MB. Effects of altering intensity during 1-MHz ultrasound treatment on increasing triceps surae temperature. *J Sport Rehabil*. 2004;13:275–286.
- Gallo JA, Draper DO, Brody LT, Fellingham GW. A comparison of human muscle temperature increases during 3-MHz continuous and pulsed ultrasound with equivalent temporal average intensities. *J Orthop Sports Phys Ther*. 2004;34:395–401.
- Merrick MA, Mihalyov MR, Roethemeier JL, Cordova ML, Ingersoll CD. A comparison of intramuscular temperatures during ultrasound treatments with coupling gel or gel pads. *J Orthop Sports Phys Ther*. 2002;32:216–220.
- Merrick MA, Jutte LS, Smith ME. Cold modalities with different thermodynamic properties produce different surface and intramuscular temperatures. *J Athl Train*. 2003;38:28–33.
- Palmer JE, Knight KL. Ankle and thigh skin surface temperature changes with repeated ice pack application. *J Athl Train*. 1996;31:319–323.
- Myrer JW, Measom GJ, Fellingham GW. Intramuscular temperature rises with topical analgesics used as coupling agents during therapeutic ultrasound. *J Athl Train*. 2001;36:20–26.
- Myrer JW, Myrer KA, Measom GJ, Fellingham GW, Evers SL. Muscle temperature is affected by overlying adipose when cryotherapy is administered. *J Athl Train*. 2001;36:32–36.
- Hopkins JT, Ingersoll CD, Edwards J, Klootwyk TE. Cryotherapy and transcutaneous electric neuromuscular stimulation decrease arthrogenic muscle inhibition of the vastus medialis after knee joint effusion. *J Athl Train*. 2002;37:25–31.
- Otte JW, Merrick MA, Ingersoll CD, Cordova ML. Subcutaneous adipose tissue thickness alters cooling time during cryotherapy. *Arch Phys Med Rehabil*. 2002;83:1501–1505.
- Garrett CL, Draper DO, Knight KL. Heat distribution in the lower leg from pulsed short-wave diathermy and ultrasound treatments. *J Athl Train*. 2000;35:50–55.
- Drust B, Atkinson G, Gregson W, French D, Binningsley D. The effects of massage on intra muscular temperature in the vastus lateralis in humans. *Int J Sports Med*. 2003;24:395–399.
- Merrick MA, Knight KL, Ingersoll CD, Potteiger JA. The effects of ice and compression wraps on intramuscular temperature at various depths. *J Athl Train*. 1993;23:236–245.
- National Institution of Standards and Technology. Standards. Available at: <http://www.nist.gov/public-affairs/standards.htm>. Accessed January 24, 2005.
- Ramsey F, Schafer D. *The Statistical Sleuth: A Course in Methods of Data Analysis*. 2nd ed. Pacific Grove, CA: Duxbury Press; 2002.
- Taylor BN, Kuyatt CE. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. 1994 ed. NIST Technical Note 1297. Gaithersburg, MD: National Institute of Standards and Technology; 1994.
- Thomas JR, Nelson JK. Measuring research variables. In: *Research Methods in Physical Activity*. 4th ed. Champaign, IL: Human Kinetics; 2001: 181–200.
- Temperature microprobes. *Physitemp Precision Temperature Specialists*. Clifton, NJ: Physitemp Instruments, Inc; 1999.
- Appendix specifications. *MMS3000-T6V4 Owner's Manual*. Christchurch, New Zealand: Commtest Instruments Limited; 2001:75.
- Ordering, terms, & warranty information. *Physitemp Precision Temperature Specialists*. Clifton, NJ: Physitemp Instruments, Inc; 1999.
- Mish FC, ed. *The Merriam-Webster's Collegiate Dictionary*. 11th ed. Springfield, MA: Merriam-Webster; 2003.
- Egan EJ, ed. *Taber's Cyclopedic Medical Dictionary*. 18th ed. Philadelphia, PA: FA Davis; 1997.
- National Institution of Standards and Technology: physics laboratory. The NIST Reference on Constants, Units, and Uncertainty: Glossary, 10/2000. Available at: <http://physics.nist.gov/cuu/Uncertainty/glossary.html>. Accessed February 21, 2005.
- Iso-Thermex Electronic Thermometer Instruction Manual*. Columbus, OH: Columbus Instruments International Corp; 1992.